Detailed Notes for Selected Variables

SUBTYPE—NDI Submission record type

The categories for the variable SUBTYPE were developed from the variables used in the selection step of the NDI matching process. Although a NDI submission record may contain data for up to 12 match items, only a subset of those (Social Security Number (SSN) and components of name and date of birth) are used to determine which NDI records will be retrieved in the selection process.

The categories for SURTYPE are sex-specific. Within sex, the order of categories reflects the probability of retrieving a correct death certificate for a now deceased survey respondent. A complete submission record has all of the following variables: SSN, date of birth (at least month and year), and complete name (first, middle initial, surname) although a blank middle initial is considered to be a valid value. For females, complete name also includes birth surname.

1 Male – Complete (SSN; First Name; Last Name; Month and Year of Birth)
2 Male – Complete except for SSN (First Name; Last Name; Month and Year of Birth)
3 Male – All other combinations
4 Female – Complete (SSN; First Name; Last Name; Birth Surname; Month and Year of Birth)
5 Female – Complete except for SSN (First Name; Last Name; Birth Surname; Month and Year of Birth)
6 Female – Complete except for Birth Surname (SSN; First Name; Last Name; Month and Year of Birth)
7 Female – Missing both SSN and Birth Surname (First Name; Last Name; Month and Year of Birth)
8 Female – All other combinations

The value of SUBTYPE corresponds to the most complete submission record for each eligible participant. Occasionally, due to supplemental information collected from other administrative record matches, some participants will have submission records both with and without an SSN. In these cases, the value of SUBTYPE reflects the most complete submission record.

If other samples with known mortality are compared to NHANES III mortality, both should include SUBTYPE as a stratification variable.

Detailed Notes for Selected Variables

POS_TOTAL1—NDI Total Possible Score

The NDI total possible score is calculated by summing the weight values for non-missing NDI submission data items, plus the average weight value for missing items. The NDI weight values for first name and marital status use sex-specific values if the item is missing. Although some weights (first name and marital status) are stratified by additional characteristics when they are present, only sex is used as a stratifier for the average value when they are missing. As blank is a valid value of middle initial, no possibility exists for it to be missing. This table shows the mean, minimum and maximum values for all NDI weights, rounded to one decimal place.

NDI submission record variable	Minimum	Maximum	Mean
Sex	0.9	1.1	1.0
Race	0.3	5.0	2.7
Day of birth	4.9	4.9	4.9
Month of birth	3.5	3.7	3.6
Year of birth	3.0	8.3	6.4
State of birth	3.8	14.0	6.8
State of Residence	3.4	17.5	7.2
SSN digits 3,6,7,8,9	3.3	3.3	3.3
SSN digit 1	2.1	12.8	5.7
SSN digit 2	3.1	3.8	3.3
SSN digit 4	2.1	6.2	3.9
SSN digit 5	2.6	4.4	3.5
Last dame	6.1	15.9	14.4
First name – male	8.8	23.3	19.3
First name - female	7.8	23.8	19.5
Middle initial – male	3.0	12.4	7.0
Middle initial – female	3.0	15.4	7.2
Marital status – male	3.0	20.8	8.6
Marital status - female	2.2	20.4	9.0

Detailed Notes for Selected Variables

CLASS—NDI Class codes

Class 1: Exact match. Agrees on at least 8 (of 9) digits of SSN, first name (including NYSIIS match), middle initial (including blank), last name (including NYSIIS match), year of birth (+/- 3 years), month of birth, sex, and state of birth

Class 2: Near exact match. Agrees on at least 7 (of 9) digits of SSN and at least 5 more of the following items: first name (including NYSIIS match); middle initial (including blank); last name (including NYSIIS match); day of birth; year of birth (+/- 3 years); race; sex; marital status; state of birth.

Class 3: Close match. There are two types of class 3 matches, Type A where SSN is missing and Type B where SSN does not agree but there is evidence that SSN has been incorrectly recorded.

> Type A: SSN is unknown on either the submission record or the NDI record, but last name matches (including NYSIIS match) and at least 7 (of 8) of the following items also agree: first name (including NYSIIS match); middle initial (including blank); last name (including NYSIIS match); day of birth; year of birth (+/- 3 years); race; sex; marital status; state of birth.
>
> Type B: SSN is known and 3 or more digits do not agree, but 8 (of 9) of the following items also agree: first name (including NYSIIS match); middle initial (including blank); last name (including NYSIIS match); day of birth; year of birth (+/- 3 years); race; sex; marital status; state of birth, specifically including last name and sex agreement. These records are initially classified as 5 matches but later switched to class 3.

Class 4: SSN is unknown on either the submission record or the NDI record, and less than 8 of the items from class 3 agree.

Class 5: Definite non-match: SSN present and fewer than 7 digits agree, and the record is not switched to Class 3 Type B. Also records where 7 or more digits of SSN agree, but less than 5 other items agree (see Class 2 above) and records where 7 or more digits of SSN and last name agree, but sex and first name do not agree (suggesting that a deceased spouse’s SSN is recorded).

Note: No class 5 records are included on the NHANES III mortality file.

Detailed Notes for Selected Variables

MATCH and *NONMATCH*-- Probability of NDI Match and Non-match

Probabilities of a match or a non-match have been estimated for each of the five NDI classes. The estimated probabilities for NDI classes 2, 3, and 4 are based on logistic regressions on known correct matches from the NHEFS training sample (see matching methodology document for a description of NHEFS training sample). Class 1 matches have an estimated probability of a match equal to 1 and non-match equal to 0, since all class 1 matches are considered exact matches. Class 5 records have an estimated probability of a match equal to 0.002 and non-match equal to 0.998, based on the very small proportion of NHEFS correct matches that are class 5 (1 out of 600 unique class 5 records). Separate models were developed for the probability of a match and the probability of a non-match. The coefficients from the models are used to compute the estimated probabilities. The match and non-match models contain indicator variables corresponding to each submission record variable. For the match model, these variables equal 1 if the NHANES III and NDI records match on that particular variable and 0 otherwise. For the non-match model, the indicator variable equals 1 if the NHANES III and NDI records do not match on that variable. The model for the probability of a match includes variables for first name, month of birth, year of birth, sex, race(white, black, other), state of birth, and class as a categorical variable with class=2 as the reference category. The model for the probability of a non-match includes variables for first name, month of birth, year of birth, sex, race, state of birth, state of residence, and class as a categorical variable with class=2 as the reference category.

The final models are:

match = -13.1702 + mm_fname*3.2717 + mm_mob*4.8066 +
mm_yob*2.1905 + mm_sex*5.2475 + mm_race*2.4247 +
mm_sob*3.2503 + class3*-6.8223 + class4*-9.3514

nonmatch = -8.0342 + nm_fname*3.2519 + nm_mob*4.6921 +
nm_yob*1.9463 + nm_sex*5.3509 + nm_race*2.4853 + nm_sob*2.9460 +
nm_sor*3.1147 + class3*6.1567 + class4*8.9210

The resulting predictions are transformed into probabilities using the inverse of the logit transform, $e^{x*b} / 1+e^{x*b}$.

The difference of the two estimated probabilities, p_match – p_nonmatch, is an estimate of the uncertainty of a match with values near zero being the most uncertain.